

Towards More Discriminative and Robust Iris Recognition by Learning Uncertain Factors

Jianze Wei, Huaibo Huang, Yunlong Wang, Ran He, *Senior Member, IEEE*, Zhenan Sun, *Member, IEEE*,

Abstract—The uncontrollable acquisition process limits the performance of iris recognition. In the acquisition process, various inevitable factors, including eyes, devices, and environment, hinder the iris recognition system from learning a discriminative identity representation. This leads to severe performance degradation. In this paper, we explore uncertain acquisition factors and propose uncertainty embedding (UE) and uncertainty-guided curriculum learning (UGCL) to mitigate the influence of acquisition factors. UE represents an iris image using a probabilistic distribution rather than a deterministic point (binary template or feature vector) that is widely adopted in iris recognition methods. Specifically, UE learns identity and uncertainty features from the input image, and encodes them as two independent components of the distribution, mean and variance. Based on this representation, an input image can be regarded as an instantiated feature sampled from the UE, and we can also generate various virtual features through sampling. UGCL is constructed by imitating the progressive learning process of newborns. Particularly, it selects virtual features to train the model in an easy-to-hard order at different training stages according to their uncertainty. In addition, an instance-level enhancement method is developed by utilizing local and global statistics to mitigate the data uncertainty from image noise and acquisition conditions in the pixel-level space. The experimental results on six benchmark iris datasets verify the effectiveness and generalization ability of the proposed method on same-sensor and cross-sensor recognition.

Index Terms—Iris recognition, cross-sensor recognition, uncertainty learning, curriculum learning.

I. INTRODUCTION

IRIS recognition is widely considered to be the most accurate biometric technology that acquires ocular images and investigates their iris patterns to verify identity. Researchers believe that iris recognition has a bright future in personal identification under surveillance scenarios. However, its recognition performance is inevitably affected by various acquisition factors, including eyes (posture, motion, occlusions), devices (optical lens, sensor, electronic control unit) [1], and environment (illumination) [2], [3]. The uncertainty associated with these acquisition factors makes it challenging to capture highly satisfying imagery, and this challenge is also known

J. Wei is with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China, and with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jianze.wei@cripac.ia.ac.cn).

H. Huang, Y. Wang, R. He and Z. Sun are with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and with the Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: huaibo.huang@cripac.ia.ac.cn; yunlong.wang@cripac.ia.ac.cn; rhe@nlpr.ia.ac.cn; znsun@nlpr.ia.ac.cn). Y. Wang is the corresponding author.

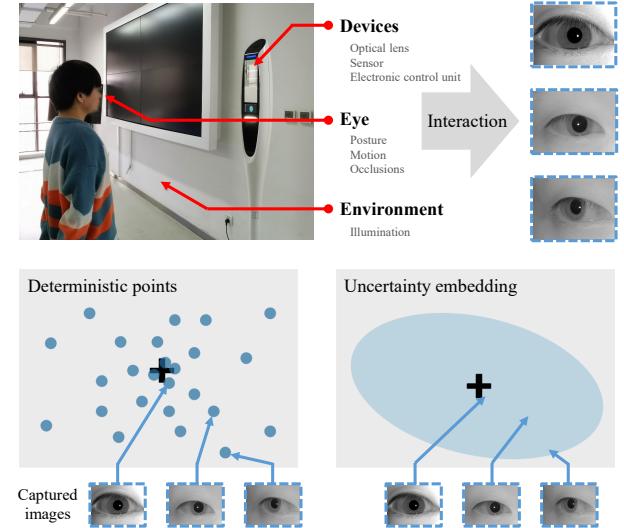


Fig. 1: The influence of uncertain acquisition factors on iris recognition. **Top:** Iris images are the interaction results of eyes, devices, and the environment. These images are inevitably affected by various acquisition factors. **Bottom-left:** Conventional deterministic embedding employs a point to represent an iris image, and this representation forces the feature extractor to learn acquisition uncertainty information (a.k.a. feature ambiguity dilemma [7]). **Bottom-right:** Uncertainty embedding adopts a distribution for feature representation to mitigate the feature ambiguity.

as data uncertainty [4]. Fig. 1 visualizes the uncertainty of the acquisition factors. It is easy to note that the data uncertainty [5], [6] makes the iris features of collected images deviate from the class center, leading to sizeable intra-class divergence.

To narrow this divergence, it is natural to pursue a powerful feature extractor to learn a compact feature representation. To the best of our knowledge, most existing iris recognition methods represent iris texture using a deterministic point. Specifically, conventional iris recognition methods [8], [9] encode the iris texture as a binary template, while deep iris recognition [10], [11] adopts a deep network to generate a feature vector. However, these deterministic point-based representation methods focus on narrowing intra-class distances of iris features but without adequately considering the data uncertainty associated with deep causal influences that lead to intra-class distances. To further reduce intra-class distances, the iris community also develops many iris-specific tricks [10], [12], such as bitshift, mask, and template binarization. It is also popular to increase the model capacity [13] of recog-

nition methods for smaller intra-class distances in a feature-level space. Even though these efforts reduce the intra-class distance at the expense of computational burden and model capacity, the inherent data uncertainty still forces deterministic point-based representation methods to learn ambiguous iris features [7] (as shown in Fig. 1). This motivates us to *skip the trap of trying to blindly narrow the intra-class distance, and instead turn to exploring data uncertainty in iris recognition.*

Recently, uncertainty learning has attracted more attention and achieved excellent success [14], [15] in multiple fields since it can improve the robustness and interpretability of deep models [16], [17]. This impressive progress encourages us to represent iris patterns as probabilistic distributions rather than deterministic points. In an ideal probabilistic distribution representation, the mean will denote the identity feature irrelevant to acquisition factors, while the variance will represent the data uncertainty associated with intra-class distances. Regardless of how the acquisition process changes, the acquisition uncertainty will only change the variance but not make the mean ambiguous.

The above analysis motivates us to design a new approach based on the data uncertainty for iris recognition. In this paper, we propose uncertainty embedding (UE) to generate a discriminative and robust iris representation and further mitigate the impact of uncertain acquisition factors in near-infrared iris recognition. We observe that the acquired images of the same class are always similar to each other during the acquisition process. Based on this observation, we make two assumptions: 1) The acquired images of the same class conform to a Gaussian distribution (simple and close to the real world) around a certain center point in the feature space; 2) The center point can be regarded as a discriminative deterministic point without uncertainty information, and it is an ideal identity feature. Based on these assumptions, we propose UE by representing the iris image using a Gaussian distribution. In UE, the mean encodes the most likely identity feature of the iris image, while the variance encodes the data uncertainty from acquisition factors. Thus, each captured image can be regarded as an instantiated iris feature sampled from the parametric probabilistic distribution with Gaussian density. UE distinguishes the identity feature from the iris feature of the imaging result affected by various acquisition factors, thus avoiding the feature ambiguity dilemma. More importantly, we can sample from UE to generate virtual features that have not been captured, thereby increasing the diversity of data available for training.

In addition, we believe that data uncertainty analysis has the potential to improve recognition performance since data uncertainty is also related to the difficulty of learning the features. Specifically, under our assumption, the iris feature with large uncertainty suggests the substantial impact of the acquisition process on imaging. It causes noticeable intra-class deviation and degrades the learning performance in the early training stage [18], [19]. Thus, dealing with these hard features with large uncertainty becomes the focus of recognition model training. Recent work [20], [21] has proven that curriculum learning can facilitate recognition accuracy by feeding features in an easy-to-hard order at different training stages. However,

it is still an open challenge to assess the feature difficulty using data uncertainty in recognition tasks. Here, we apply data uncertainty to score the difficulties of virtual features and propose uncertainty-guided curriculum learning (UGCL) to select features for optimization in an easy-to-hard order at different training stages. With this learning strategy, more hard features would be gradually fed into the model as the iteration progresses, improving the recognition performance.

Recently published literature [10] suggests that image enhancement provides a significant improvement in iris recognition performance, especially cross-database (introduced in Appendix-B) recognition with unknown acquisition conditions. Inspired by this discovery, we elaborately design an instance-level enhancement method to alleviate the data uncertainty problem. For an iris recognition system, uncertainty from noise and unknown acquisition conditions frequently occurs in the cross-database scene [22] and usually threatens its recognition performance significantly [23]. To address this problem, the proposed enhancement method modifies each normalized iris image according to its local and global statistics. With the help of this enhancement method, we alleviate the pixel-level uncertainty caused by image noise and acquisition factors.

The contributions of this paper are summarized as follows:

- We propose UE based on a Gaussian distribution to represent an iris image rather than a deterministic point. Compared with the deterministic point-based representation, UE encodes identity and uncertainty into independent components of the distribution, thus reducing the impact of the uncertain factors on identity features. In addition, we can generate virtual features by sampling from UE to increase the data diversity for training.
- We propose UGCL to optimize the model in an easy-to-hard manner for higher recognition accuracy. Acquisition factors and sampling operation entrust virtual features with different uncertainties. UGCL utilizes uncertainty as an indicator and preferentially selects easy features with small uncertainty to learn the model. With the increase in iterations, features with large uncertainty are provided more chances to train the model.
- Based on the exploration of acquisition uncertainty, we develop an instance-level enhancement module to reduce the uncertainty from image noise and unknown acquisition conditions. Cross-database recognition is a challenging task since its acquisition condition is unavailable. The developed enhancement module mitigates the acquisition uncertainty problem from local and global perspectives.
- We conduct experiments on four same-sensor and two cross-sensor benchmark iris datasets. Our method shows superior performance than the competitive methods on these datasets. This verifies the effectiveness and generalization ability of the proposed method in multiple iris recognition tasks. Models, evaluation protocols, and codes will be made freely available at <https://github.com/reborn20200813/uncertainty>.

The rest of this paper is organized as follows. In Section II, we provide a brief review of related works, especially methods

for specific domains or tasks. Section III presents our proposed method in detail. In Section IV, the adopted datasets and experimental evaluations are presented. Finally, the conclusion is given in Section V. In addition, the supplementary material provides more details about the network architecture and experimental settings.

II. RELATED WORK

A. Deep iris recognition

Compared with face recognition, iris recognition generally pays more attention to texture information instead of geometric information [24]. The preference difference between face and iris patterns makes conventional iris recognition adopt feature templates [25] containing more spatial information than feature vectors. At the same time, some techniques for conventional iris recognition [9], [8], [26], such as bitshift, mask, and template binarization, are also employed to help feature templates avoid interference from acquisition factors.

In the era of deep learning [27], it has become a trend to utilize deep networks to revolutionize iris recognition. Moreover, recently proposed approaches [10] explore the possibility of introducing deep learning into conventional iris recognition. UniNet [10] retains the typical techniques for iris recognition and introduces a deep learning model as a substitute for the filter bank of conventional iris recognition. Even if deep face recognition has made significant progress [28], [29] in recent years and received appreciation from the market and academia, a considerable number of people still doubt the feasibility of representing iris images using feature vectors. In fact, some pioneering works have explored this question.

For same-sensor iris recognition, [11] employs maxout-feature-map (MFM) [30] as the activation function and generates more compact and discriminative feature vectors. [31] adopts an off-the-shelf CNN model to replace the Gabor filter and binary template in conventional iris recognition. [32] first borrows the idea of network architecture search (NAS) and builds the criteria considering the model size and computation to search the best network architecture automatically for iris recognition. Dynamic graph representation (DGR) [33] first applies deep graphical models for biometrics to overcome occlusion situations. For cross-sensor iris recognition, Deepirisnet [13] first explores this task using a deep network with batch normalization (BN) layers [34] and the fine-tuning. CSIN [35] narrows the distribution gap by utilizing sensor-specific information and adversarial strategy.

B. Deep uncertainty learning

In recent years, deep uncertainty learning has attracted more attention from academia and industry. It provides a reliability assessment decision to address various tasks in computer vision and natural language processing.

Deep uncertainty learning can be roughly categorized into two major types: data uncertainty (also known as aleatoric uncertainty) [4] and model uncertainty (also known as epistemic uncertainty) [36]. Data uncertainty focuses on the inherent noise in the given training data and captures the uncertainty

from data noise. Unlike data uncertainty, model uncertainty comes from the ignorance of model parameter noise.

The recently published literature about uncertainty shows exciting progress in model robustness and interpretability in low- and high-level computer vision tasks. For semantic segmentation, Bayesian SegNet [14] first utilizes Monte Carlo sampling with dropout to measure model uncertainty for each class and further predicts pixel-wise class labels. For object detection, Gaussian YOLOv3 [15] redesigns the loss function by introducing the Gaussian function and predicts the localization uncertainty to indicate the reliability of the bounding box.

In addition to the exciting progress in these low-level tasks, deep uncertainty learning also shows similar success in high-level computer vision tasks. For person Re-ID, DistributionNet [6] modeled feature uncertainty explicitly using a Gaussian distribution to address both the noisy label and outlying sample problems. For face recognition, [37], [38] utilized Bayesian estimation to obtain the uncertainty information of samples to refine the decision boundary. [7] and [39] considered data uncertainty in the face recognition task and re-defined the similarity metric based on the Gaussian distribution. Compared with previous deep uncertainty learning methods, the proposed method utilizes uncertainty (variance) to refine the classification loss and guide the model training.

III. METHOD

As discussed earlier, the proposed method estimates the acquisition uncertainty for iris images. With the estimated uncertainty, our method learns a discriminative representation based on probabilistic distributions and builds a smart training strategy guided by uncertainty. The overall framework is illustrated in Fig. 2. Section III-A to Section III-C describe UE, UGCL and the enhancement module in detail. Section III-D details the rationale of the proposed method. Appendix A provides the architecture of the encoder.

A. Uncertainty embedding

Due to the impact of acquisition factors, the acquired iris image naturally contains data uncertainty even if strict requirements are obeyed during the acquisition stage. If we visualize the iris images of the same eye in a high-level semantic space using deterministic points, these points are clustered around a deterministic center, as shown in Fig. 1. The deterministic center represents an ideal identity feature that only contains personal identifiable information, while the deterministic points around the center are the iris features of the captured imaging results affected by uncertain factors. Training with these deterministic points usually makes the model learn nonidentity information and fall into the feature ambiguity dilemma [7]. This motivates us to develop a new representation to describe the image.

To represent the identity and uncertainty simultaneously, we accordingly adopt a multivariate Gaussian distribution. In particular, the n -th iris image $\mathbf{X}_n^o \in \mathbb{R}^{H \times W}$ with a resolution of $H \times W$ can be extracted as an uncertainty embedding

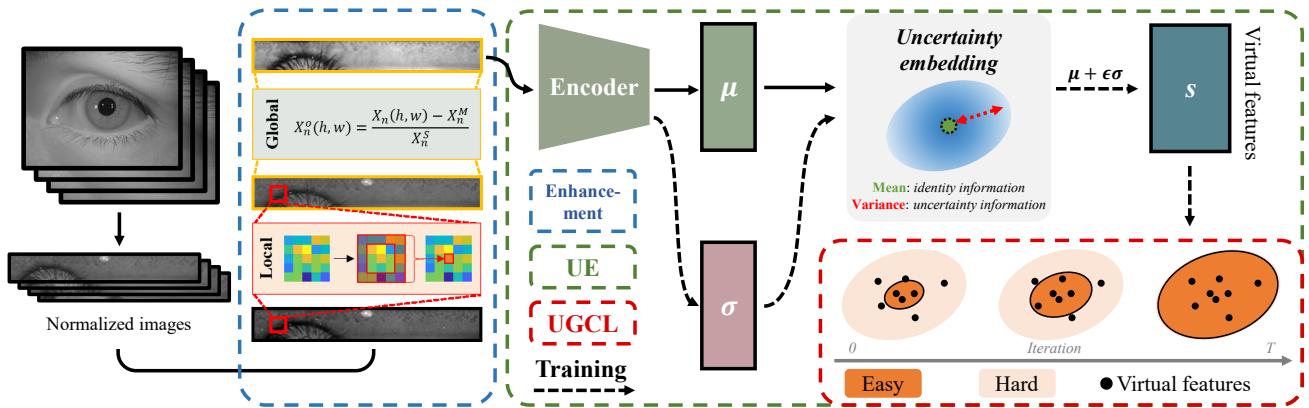


Fig. 2: An illustration of the proposed model architecture. For the normalized iris images, the enhancement module alleviates the data uncertainty and enhances the iris images from the local and global statistical perspectives. Unlike previous deterministic point-based representations, Uncertainty embedding (UE) utilizes a probabilistic distribution to describe the iris images in a high-level semantic space. The variation σ encodes the uncertainty of the input image. Moreover, uncertainty-guided curriculum learning (UGCL) generates the difficulty score for each virtual feature according to the variation σ and random noise ϵ and this guides the model learning in an easy-to-hard order. The dotted line is only available in the training stage.

representation z_n in the feature space, and z_n can be defined using a Gaussian distribution as follows:

$$p(z_n | \mathbf{X}_n^o) = \mathcal{N}(z_n; \boldsymbol{\mu}_n, \boldsymbol{\sigma}_n^2 \mathbf{I}), \quad (1)$$

where the key parameters of the Gaussian distribution, $\boldsymbol{\mu}_n \in \mathbb{R}^D$ and $\boldsymbol{\sigma}_n \in \mathbb{R}^D$, are predicted by an encoder (the detailed architecture is illustrated in Appendix A). $\boldsymbol{\mu}_n$ represents the identity feature generated by the μ -branch, while $\boldsymbol{\sigma}_n$ encodes the data uncertainty along each feature dimension and is generated by the σ -branch. D is the feature length. \mathbf{I} is an identity matrix. The iris feature of the input image can be regarded as a special case of stochastic points sampled from $\mathcal{N}(z_n; \boldsymbol{\mu}_n, \boldsymbol{\sigma}_n^2 \mathbf{I})$.

For the parameter learning of the Gaussian distribution, we apply the re-parameterization technique [40] here. Specifically, it generates virtual features s_n by sampling random noise ϵ_n from a normal distribution, i.e., $s_n = \boldsymbol{\mu}_n + \epsilon_n \otimes \boldsymbol{\sigma}_n$, where $\epsilon_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the random noise and \otimes denotes the Hadamard product. Then, the virtual features are used to optimize the objective function. The re-parameterization trick lets the gradient backpropagate through the sampling operation as usual. In the testing stage, ϵ_n , which represents uncertainty information, is an unnecessary component since it does not contribute to the performance improvement in iris recognition. As shown in Fig. 2, a dotted line is employed in the training stage, and it is unavailable in feature matching. In our implementation, we adopt cosine similarity (more details in Appendix-C.B) to estimate the matching score.

In addition, it should be noted that training with only a classification loss (Eq. 6) easily makes the model fall into a naïve solution, i.e., the model tends to predict a smaller σ with decreasing classification loss. Thus, the constraint on σ is necessary to avoid the naïve solution. Here, a Kullback-Leibler (KL) regularization term is adopted for the constraint on σ . The KL regularization term forces the predicted distribution $\mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\sigma}_n^2 \mathbf{I})$ to be close to a normal distribution, $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

This regularization can be written as

$$\begin{aligned} \mathcal{L}_{kl} &= \mathbb{E}\{KL[\mathcal{N}(z|\boldsymbol{\mu}, \boldsymbol{\sigma}) || \mathcal{N}(\epsilon|\mathbf{0}, \mathbf{1})]\} \\ &= -\frac{1}{2N} \sum_{n=1}^N \sum_{d=1}^D (1 + \log \sigma_{n,d}^2 - \mu_{n,d}^2 - \sigma_{n,d}^2). \end{aligned} \quad (2)$$

B. Uncertainty-guided curriculum learning

As mentioned above, the acquired image under strict requirements still suffers from different degrees of data uncertainty. A similar situation also occurs in virtual features due to the estimated variance and sampled noise. The divergence in data uncertainty suggests that different features would not provide the same benefit on optimization, i.e., some training features are hard for the model while others are not. The progress in CL [20] directly points out that hard features play a negative role in the convergence at the early training stage, and this negative role would even degrade the final performance. Therefore, compared to training without any strategy, an easy-to-hard learning strategy helps the recognition model obtain higher accuracy and better generalization.

To introduce curriculum learning into deep iris recognition based on data uncertainty, we have to address the key point of CL, namely, assessment of difficulties of the virtual features. According to the definition of virtual features, how much the virtual feature is affected by uncertainty information (i.e., the distance between the virtual feature and the identity feature) is driven by estimated variance σ_n and random noise ϵ_n . Here, we compute the harmonic mean of the Hadamard product of σ_n and ϵ_n as the difficulty score to design the curriculum. The difficulty score of a virtual feature sampled from the n -th uncertainty embedding, r_n , is obtained by the following formula:

$$r_n = \frac{D}{\sum_{d=1}^D \frac{1}{\sigma_{n,d} \epsilon_{n,d}}}, \quad (3)$$

where $\sigma_{n,d}$ and $\epsilon_{n,d}$ are the d -th components of $\boldsymbol{\sigma}_n$ and $\boldsymbol{\epsilon}_n$, respectively. According to Eq. 3, the difficulty score of the virtual feature is determined by both $\boldsymbol{\sigma}_n$ and $\boldsymbol{\epsilon}_n$.

With the difficulty scores, it is easy to sort the virtual features in an easy-to-hard order, where a lower score usually corresponds to an easier feature. Then, the sorted features are divided into two pools of easy and hard features according to a pool-parameter α . α is a dynamic threshold for the division of easy and hard features. Virtual features with scores below the threshold are arranged into the easy feature pool; otherwise, they are regarded as hard features. Here, this division process is described as a sampling function formulated as

$$p_n = \begin{cases} 1 & \text{if } r_n < \alpha \\ 0 & \text{if } r_n \geq \alpha \end{cases} \quad (4)$$

At the training iteration, the easy-sample pool without hard features is adopted first to train the model. Then, we adjust α to put more hard features with large difficulty scores from the hard feature pool into the easy feature pool. This uncertainty-based learning strategy improves the learning and recognition performance of the model. For a smoother adjustment of α , a sigmoid function is employed to control the expansion of the easy feature pool. The sigmoid function can be written as

$$\alpha = \frac{1}{1 + e^{-t}}, \quad (5)$$

where t denotes the training epochs.

Based on the sampling function (Eq. 4), we can define a classifier with the weight $\mathbf{W} \in \mathbb{R}^{D \times C}$ (C is the number of classes) and obtain an uncertainty-guided softmax loss after instance sampling. Taking ArcFace loss [41] as an example, the uncertainty-guided version can be rewritten as

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_{n=1}^N p_n \log \frac{e^{g(\cos(\theta_{y_n} + m))}}{e^{g(\cos(\theta_{y_n} + m))} + \sum_{c=1, c \neq y_n}^C e^{g \cos \theta_c}}, \quad (6)$$

where $\cos \theta_c = \frac{\mathbf{W}_c^T \mathbf{s}_n}{\|\mathbf{W}_c\| \|\mathbf{s}_n\|}$ measures the cosine distance between \mathbf{s}_n and \mathbf{W}_c (the c -th column vector). m denotes the margin penalty between \mathbf{s}_n and \mathbf{W}_c , which enhances the intra-class compactness and inter-class discrepancy. g is the feature scale. We follow [41] and set $g = 64$ and $m = 0.5$. UGCL provides a weighting scheme to compute loss, and any other softmax losses can be used here as an alternative to ArcFace loss.

Thus, the final loss can be rewritten as

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{kl}, \quad (7)$$

where λ is the trade-off parameter, and it balances classification loss and KL regularization term.

Conventional CL requires an extra forward propagation stage to generate the difficulty score for each sample before each epoch, which significantly increases the computational burden. More importantly, the proposed method cannot estimate the difficulty of the features before each epoch due to random noise ϵ . To address this problem, a bolder and more radical solution comes to our mind for the efficiency improvement of UGCL. In our solution, we scale the division of the easy and hard feature pools down from dataset level to batch level, i.e., for virtual features in the same batch, the features with lower scores are arranged into the easy feature pool, and the others are arranged into the hard feature pool. In

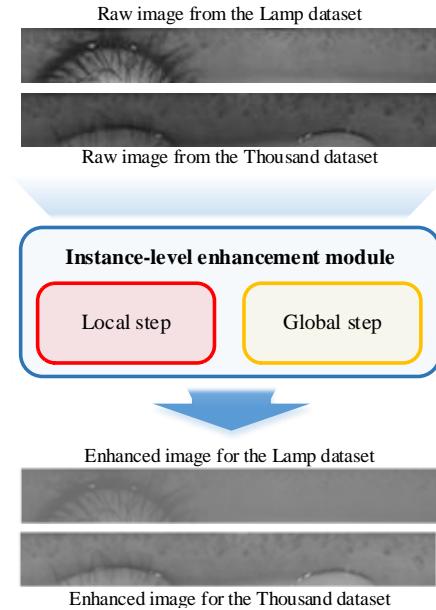


Fig. 3: Illustration of the instance-level enhancement module. Compared with the raw images, these enhanced images look more alike as they are captured in the same acquisition scene.

this way, it is not necessary to add extra forwards to generate difficulty scores for features. This significantly shortens the training time.

C. Image enhancement module

Since it is always unknown to identify objects and the acquisition process in the actual recognition scenario, cross-database recognition becomes an inevitable task for any biometric technology. For iris recognition, there is a significant imaging gap [18] between training and testing sets in the cross-database scenario. As mentioned in Section I, this imaging gap is usually caused by uncertainty from the unknown acquisition conditions and image noise. Based on the above analysis, we propose an instance-level enhancement module to mitigate the uncertainty from local and global statistical perspectives, as shown in Fig. 2.

Image noise is an inevitable challenge for any computer vision task, and it also heavily affects the accuracy performance of iris recognition methods. The enhancement module focuses on the local region and computes the median of the region as the pixel of the center position. This local non-linear smoothing operation alleviates the uncertainty from image noise from a local perspective. Here, we adopt a sliding window of 3×3 to avoid over-smoothing.

Due to the unrepeatability of the acquisition process, each image has its particular brightness and contrast [42], and the variation in brightness and contrast is more evident in the cross-sensor scenario [43]. Regarding global uncertainty, the enhancement module corrects the image based on the first-order and second-order statistical information of the entire normalized image. Specifically, assuming that $\mathbf{X}_n \in \mathbb{R}^{H \times W}$ denotes the n -th normalized image with a width of H and

height of W , we can obtain the enhanced image after statistical enhancement using the following formula:

$$X_n^o(h, w) = \frac{X_n(h, w) - X_n^M}{X_n^S}, \quad (8)$$

where $X_n^M = \frac{\sum_{h,w} X_n(h,w)}{H \times W}$ and $X_n^S = \sqrt{\frac{\sum_{h,w} [X_n(h,w) - X_n^M]^2}{H \times W}}$. h and w indicate the positions in the horizontal and vertical directions, respectively.

The proposed enhancement module utilizes local and global statistical operations to enhance the normalized iris image, which significantly alleviates the data uncertainty problem in cross-database recognition. In fact, we find that this module is also beneficial for handling uncertainty in the within-database setting. Fig. 3 shows some examples of the enhanced iris images. Compared with the original images, these enhanced images look more like they were captured in the same acquisition scene.

D. Method reasonability

In the acquisition process, each iris image is a single sampling result of an iris texture under a certain acquisition condition. Therefore, iris images are inevitably affected by various acquisition factors. Some of them are driven by the acquisition condition and can be regarded as certain factors. The others are out of human control, a.k.a., uncertain factors, leading to acquisition uncertainty. Regardless of certain or uncertain factors, they are not only tightly entangled with identity features but also entangled with each other.

Uncertain factors play a negative role in learning identity representation. They shift iris features extracted by the model away from the identity feature and distribute them around the identity features. Considering the shifting effect of uncertain factors, UE regards the identity feature as the center (mean) of a Gaussian distribution and iris features as the sampled point around the distribution center. To indicate the potential shift of the unknown testing image, UE encodes acquisition uncertainty as the variance of the Gaussian distribution. In other words, UE no longer regards the iris feature as the identity feature but a combination of the identity feature and uncertainty feature. The change enables the feature extractor to match image pairs using the identity feature at the cluster center rather than the iris feature shifted by uncertain factors.

UGCL also reduces the impact of the acquisition uncertainty by feeding input samples in an easy-to-hard order. Due to the randomness of the re-parameterization, virtual features could be located at an arbitrary position of the class cluster. According to our empirical experiments, training with features in the vicinity of cluster center (easy samples) can significantly improve recognition accuracy, while training with features at the cluster boundary (hard samples) prefers enhancing the robustness of the feature representation. However, training from scratch with difficult samples may degenerate the recognition performance of the model [20]. To integrate the merits from both easy and hard samples, UGCL filters out the difficult virtual samples in the early training stage to improve the recognition accuracy. This instance selection is equivalent to reducing the impact of uncertainty of iris features in the

early training stage. Then UGCL progressively introduces hard samples to improve the robustness of the feature extractor.

As for certain factors, they would be converted to uncertain ones in some scenarios, such as the cross-dataset setting. UE is unable to learn and alleviate this acquisition uncertainty since the testing acquisition condition is unavailable in the cross-dataset scenario. To mitigate the acquisition uncertainty from the unknown acquisition condition and image noise, the specific enhancement module modifies each normalized iris image from the perspectives of local and global statistics. The local step smooths the pixels according to local statistics to alleviate uncertainty from image noise, while the global step attenuates the impact of unknown acquisition factors by adjusting the first- and second-order statistics of the entire normalized images. With these steps, the enhanced images from different datasets are more intrinsically similar in a unified dataset style, and it seems that they are captured in the same acquisition scenario. The unified style contributes to the performance improvement in both within-database and cross-database recognition scenarios.

IV. EXPERIMENT

This section applies the proposed approach for near-infrared iris recognition tasks, including same-sensor (Section IV-D) and cross-sensor recognition (Section IV-E). In addition, we evaluate the proposed method under two recognition protocols, i.e., within- and cross-database protocols. The significant difference between these two protocols is whether the acquisition condition is available. More details about these two protocols are provided in Appendix-B. The subsequent sections detail the datasets, experimental settings, and reproducible results.

A. Dataset

1) *CASIA-iris-V4*: To promote research on long-range and large-scale iris recognition systems, Chinese Academy of Sciences' Institute of Automation (CASIA) constructed CASIA Iris Image Database V4.0 (or CASIA-IrisV4 for short) and released this database to the public domain [44]. Considering unpredictable acquisition variations in real-world applications, the database purposefully builds six datasets with particular topics, such as long-distance, large-scale, lamp, etc.

2) *CASIA-iris-Mobile*: Based on the mobile module produced by IrisKing, CASIA built the largest known dataset for mobile iris recognition [11]. Its difficulty lies in distance changes, eyeglasses, specular reflections, defocus and so on.

3) *ND-CrossSensor-Iris-2013 dataset*: To spur the research of cross-sensor iris recognition, the University of Notre Dame constructed this publicly available dataset for cross-sensor iris recognition [45]. The images from this dataset are collected by two iris acquisition devices, LG2200 and LG4000. These two devices are different in the location of illumination, field of view, and camera type [2].

4) *CASIA Cross Sensor Iris Recognition dataset*: CASIA Cross Sensor Iris Recognition dataset has been collected by CASIA and is a benchmark dataset for cross-sensor iris recognition [42]. In this dataset, two devices from different companies are used for acquisition, namely, IKEMB-200

TABLE I: Detailed information of six iris datasets. For the training set, the numbers of classes and images are detailed. For the testing set, we counted non-overlapping genuine (intra-class) and imposter (inter-class) pairs.

Dataset	Training set		Testing set	
	#class	#image	#Genuine	#Imposter
Same-sensor	Lamp	410	8,114	76,010
	Thousand	1,000	10,000	45,000
	Distance	142	1,799	9,773
	Mobile	180	1,800	8,100
Cross-sensor	NDCS	676	27,979	16,865
	CSIR	200	7,922	238,064
142,598,482				

produced by IrisKing and EyeGuard AD100 produced by IrisGuard.

B. Image preprocessing and Experiment protocol

For a fair comparison of recognition performance, the same localization and normalization techniques¹ are employed for the proposed method and the compared methods. Specifically, [46], [47] use the Hough circle detection to obtain the iris and pupil localization of an eye image. [48] maps an iris texture with an irregular ring shape into a rectangular normalized iris texture using an affine transform. In this paper, all iris images are normalized as rectangular images with a size of 66×540 . Since different models have various input size requirements, we resize normalized images to input images of 64×512 for UniNet, 128×256 for DGR, and 128×128 for Maxout-based models (Maxout, Maxout-BA, and ours) using bicubic interpolation.

For the within-database recognition experiments on these six datasets, we apply the official protocols as much as possible. Specifically, the experiments on all the same-sensor datasets (Lamp, Thousand, Distance, Mobile) and the CSIR dataset adopt the official division. As for the NDCS dataset, we borrow the idea of the protocol of [10] to design our protocols, i.e., left-eye images from all the subjects are used to generate the training set, and the testing set consists of all right-eye images of the subjects. For cross-database recognition, we only leave the testing sets since there is no demand for training images. Detailed information on the benchmark datasets is reported in Table I. In addition, for the experiments on the NDCS dataset, we apply the first five images of each class for testing considering its class imbalance.

C. Baseline methods and parameter setting

In the paper, we evaluate the recognition performance under two types of scenarios, i.e., same-sensor and cross-sensor recognitions. For same-sensor recognition, we compare the proposed method with several benchmark methods for iris recognition, including Maxout [11], Maxout-BA (the baseline model), DGR [33], log-Gabor coding (Gabor for short) [25], OM coding (OM for short) [9], UniNet [10]. Maxout utilizes the maxout-feature-map (MFM) as the activation function and designs a lightweight deep model for iris and periocular recognition. Maxout-BA upgrades Maxout by introducing BN and

¹CASIA provides normalized images, and we re-implement the experiments using these images. These images can be downloaded from <http://www.cripacsir.cn/dataset/>.

the recently proposed ArcFace loss [41]. It is a key baseline method (marked with †) for our experiment. DGR introduces a graph neural network into the deep iris recognition domain and extracts distortion-invariant keypoints as additional information for identification. All of them are deep learning models for iris recognition under the within-database protocol. As for log-Gabor coding and OM coding, both methods adopt pre-defined handcrafted parameters to design the bank of filters. In UniNet, the pre-trained parameters have been provided by the authors for the multi-branch deep model. Thus, these three methods are additionally employed for evaluation comparison under the cross-database protocol due to their pre-defined parameters. For cross-sensor recognition, we adopt CSIN [35], a deep model for the cross-sensor task, instead of DGR for performance comparison. Then Maxout, Maxout-BA, and CSIN are utilized as comparison methods under the within-database protocol, while log-Gabor coding, OM coding, Maxout, and Maxout-BA are employed for recognition comparison under the cross-database protocol. We reimplement these methods and assume that these results are their best performance.

In the experiments, we follow hyperparameter settings of the published literature and adjust several hyperparameters to uniform values. Specifically, we adopt a claimed setting for hyperparameters in UniNet, CSIN, and DGR. For Maxout-based methods, we adopt a feature length of 256 and batch size of 300. More details for pre-training and optimization are provided in Appendix-C. For conventional iris recognition methods, both log-Gabor coding and OM coding generate the templates with a code length of 1024 bytes and without multi-scale. For log-Gabor coding, we set the wave-length to 18 and the standard deviation to $0.5 \times$ wavelength. The bitshift operation shifts the pixel in the range of [-8, 8]. For OM coding, we select the two-lobe model with an inter-lobe distance of 7 and orientation of 0° . The distance is obtained by comparing the template with 11 filtered results with different orientations in the testing stage.

D. Same-sensor matching

The same-sensor iris recognition is the benchmark task under the near-infrared light scenario for recognition evaluation. To verify the effectiveness of the proposed method, we conduct the same-sensor recognition experiment on four benchmark datasets, CASIA-v4-Lamp (Lamp), CASIA-v4-Thousand (Thousand), CASIA-v4-Distance (Distance), and CASIA-Mobile-S3 (Mobile). The results under the within-database protocol are listed in Table II, while the results under the cross-database protocol are reported in Table III. In addition, Fig. 4 and Fig. 5 plot DET curves of different approaches for a qualitative comparison.

Results under the within-database protocol. Table II illustrates the results under the within-database protocol. According to the results in Table II, it is easy to reach the following three conclusions.

1. Our method achieves better and comparable results than the compared methods. Specifically, our method has the lowest EER of 0.44% on Lamp, 0.63% on Thousand, 1.09% on Distance, and 0.42% on Mobile. More concretely, compared

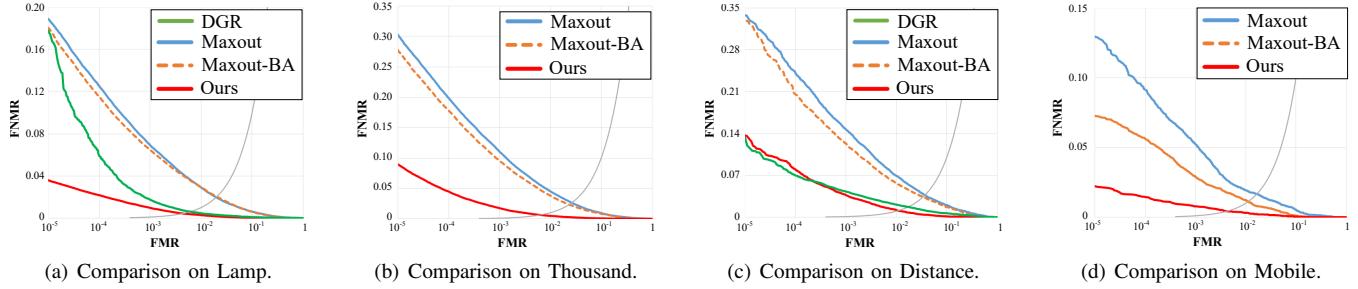


Fig. 4: DET curves of different methods on four **same-sensor** iris datasets under the **within-database** protocol. The red solid lines are the results of the proposed method. (best viewed in color)

TABLE II: Quantitative comparison on four **same-sensor** datasets under the **within-database** protocol (%). Note: \dagger means the baseline method.

	FNMR@FMR	Maxout	DGR	Maxout-BA \dagger	Ours
Lamp	EER	2.09	0.61	2.02	0.44 (\downarrow 1.58)
	10^1	0.61	0.09	0.59	0.02
	10^{-1}	7.95	1.74	6.45	0.97
	10^{-3}	21.18	17.38	18.05	3.60
Thousand	EER	2.90	–	2.33	0.63 (\downarrow 1.70)
	10^1	1.08	–	0.82	0.06
	10^{-1}	12.21	–	9.55	1.74
	10^{-3}	33.04	–	27.79	8.95
Distance	EER	3.66	1.71	3.41	1.09 (\downarrow 2.32)
	10^1	2.02	0.65	1.68	0.12
	10^{-1}	14.69	4.28	12.28	3.70
	10^{-3}	33.75	13.32	32.88	13.62
Mobile	EER	1.67	–	1.11	0.42 (\downarrow 0.69)
	10^1	0.54	–	0.11	0.04
	10^{-1}	5.22	–	2.86	0.65
	10^{-3}	12.91	–	7.25	2.67

with that of DGR (the state-of-the-art approach proposed in 2020), our method's EER value drops by 27.86% (Lamp) and 36.25% (Distance) on relative values, respectively. It should also be noted that DGR adopts a dynamic graph match, leading to more execution time than ours (mentioned in Section IV-H). Thus, the proposed approach achieves a better recognition performance and decreases the execution time simultaneously compared with DGR.

2. *Our method shows a significant improvement over the baseline model.* By introducing BN and Arcface loss into Maxout [11], we obtain a new benchmark model with high accuracy, namely, Maxout-BA. It shows a slight improvement over Maxout on recognition performance. Compared with the baseline model, our method has a significant drop in all error indices, i.e., EER, FNMR@FMR={ $10^{-1}, 10^{-3}, 10^{-5}$ }. Specifically, taking EER as an example, the proposed method drops by 1.58% (Lamp), 1.70% (Thousand), 2.32% (Distance), and 0.69% (Mobile) on absolute values compared with baseline, respectively. Compared to BN and Arcface loss, the performance improvement contributed by UE is more obvious. It demonstrates the effectiveness of the proposed method on same-sensor recognition under the within-database protocol.

3. *The superior performance of our method verifies the feasibility of mitigating the uncertainty in iris recognition.* Since these datasets adopt distinct acquisition conditions, the acqui-

sition divergence induces significantly different uncertainties among these datasets. For example, the uncertainty of the Lamp dataset is mainly caused by variations in illumination, while the uncertainty of the Distance dataset comes from the variations in acquisition distance. According to Table II, the proposed method achieves significant improvements over the baseline model on these datasets with different uncertainties. The performance improvement on multiple datasets shows that it is feasible to mitigate the uncertainty in iris recognition.

Results under the cross-database protocol. For same-sensor iris recognition under the cross-database protocol, Table III reports the detailed results of our method and the compared methods. According to Table III, we can reach similar conclusions as those under the within-database protocol.

1. *The proposed method shows a better ability to cross-database iris recognition than compared methods on these four datasets.* Our method has the lowest EER of 0.45% on Lamp, 0.98% on Thousand, 1.17% on Distance, 0.53% on Mobile. These results outperform the three benchmark methods (log-Gabor coding, OM coding, and UniNet) by a large margin. The performance advantage in cross-database scenarios further verifies the proposed method's generalization ability, even when both objects and acquisition conditions are inconsistent.

2. *Compared with the baseline model, our method shows a significant improvement in recognition performance.* The EER values of our method drop by 1.25% (Lamp), 1.76% (Thousand), 2.55% (Distance), and 0.94% (Mobile) on absolute values compared with baseline, respectively. In the cross-database setting, the unknown acquisition conditions usually cause more severe challenges in imaging and feature representation. The proposed method with our enhancement module properly handles these challenges. Compared with the results under the within-database protocol, our enhancement module provides more significant performance improvement for unknown objects and acquisition conditions.

3. *Our enhancement module can effectively reduce the impact of the data uncertainty from the cross-database scenario.* According to the results in Table II and Table III, we can obtain the average EER gap between the results under two protocols using the following formula:

$$\text{average EER gap} = \frac{\sum_k |\text{EER}_k^{cross} - \text{EER}_k^{within}|}{K}, \quad (9)$$

where EER_k^{cross} and EER_k^{within} denote the EER values on the k -th dataset under cross-database and within-database protocols, respectively. K represents the number of datasets; here,

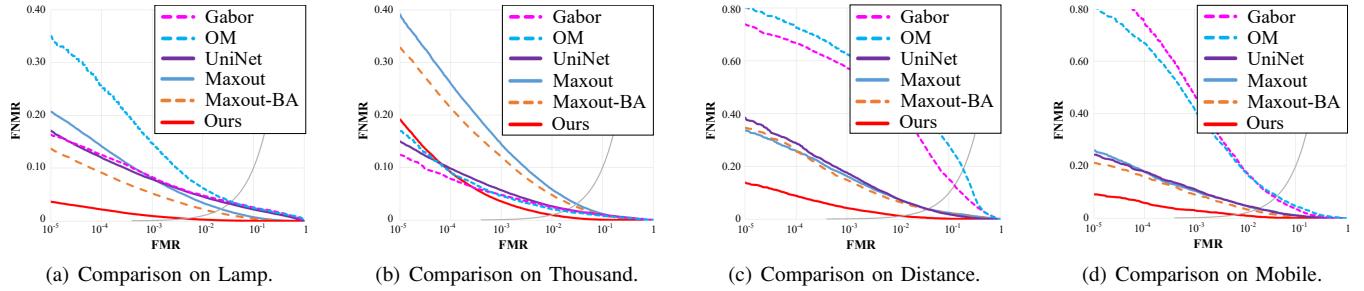


Fig. 5: DET curves of different methods on four **same-sensor** iris datasets under the **cross-database** protocol. The red solid lines are the results of the proposed method. (best viewed in color)

TABLE III: Quantitative comparison on the Lamp, Thousand, Distance, and Mobile datasets under the **cross-database** protocol (%). Note: † means the baseline method.

FNMR@FMR		Gabor	OM	UniNet	Maxout	Maxout-BA†	Ours
Lamp	EER	3.56	3.65	3.30	2.28	1.70	0.45 ($\downarrow 1.25$)
	10^1	2.60	2.47	2.20	0.75	0.44	0.03
	10^{-1}	8.34	14.65	7.92	8.10	5.25	0.95
	10^{-3}	16.30	34.95	17.12	20.72	13.68	3.68
Thousand	EER	1.96	1.75	2.02	3.08	2.74	0.98 ($\downarrow 1.76$)
	10^1	1.02	0.82	1.04	1.20	0.98	0.08
	10^{-1}	4.81	4.62	5.71	14.47	12.04	3.57
	10^{-3}	12.42	16.94	14.99	39.18	32.87	19.18
Distance	EER	12.79	18.97	3.76	3.79	3.72	1.17 ($\downarrow 2.55$)
	10^1	15.09	27.81	1.81	2.38	2.14	0.13
	10^{-1}	57.38	62.57	17.55	16.24	14.83	4.16
	10^{-3}	73.61	80.27	38.08	33.54	34.48	13.54
Mobile	EER	3.93	4.27	1.86	1.78	1.47	0.53 ($\downarrow 0.94$)
	10^1	1.74	2.32	0.51	0.58	0.19	0.05
	10^{-1}	23.42	20.69	5.50	5.25	4.51	1.48
	10^{-3}	46.56	40.27	12.15	12.94	10.56	4.53

$K=4$. Compared with Maxout and Maxout-BA, our method has the smallest average EER gap of 0.13% (Maxout: 0.15%, Maxout-BA: 0.35%). Intuitively, the enhancement module with statistical enhancement could roughly unify the brightness and contrast of input images according to the local and global statistics of the image's pixel values. This unification reduces the influence of data uncertainty under the cross-database recognition scenario, which suggests that it is feasible to employ deep iris recognition in a cross-database recognition scenario.

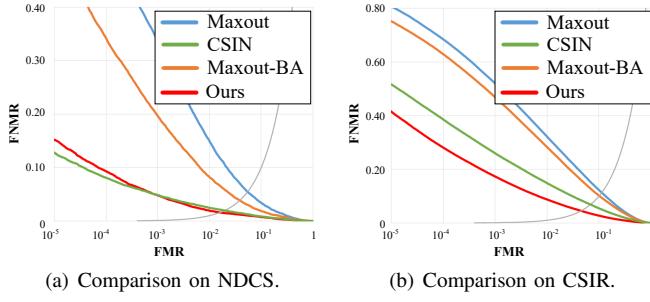
4. It is feasible to distribute deep iris recognition in the recognition system. As a representative data-driven method, deep learning discards the previous handcrafted parameter setting and automatically learns a bank of large-scale parameters for feature extraction. Under the cross-database protocol, deep iris recognition shows competitive results in comparison to conventional methods for same-sensor iris recognition. In addition to the advantage of high accuracy, deep iris recognition also enjoys two other advantages over conventional methods. 1) Looser requirement of professional knowledge and time on the handcrafted tuning of parameters. 2) There is no demand for conventional iris-specific techniques, such as mask, bitshift, and template binarization. However, we should not ignore the shortcoming of deep iris recognition, i.e., the great demand

for storage for model parameters due to insufficient prior knowledge.

5. It is easier to alleviate the data uncertainty from illumination variation than other factors according to the EER gap. It should be noted that the EER gap on Lamp is the smallest gap for our method (Lamp: 0.01%, Thousand: 0.35%, Distance: 0.08%, Mobile: 0.11%). This indicates that the Lamp dataset's acquisition variation (illumination) occurs widely in iris datasets. This is consistent with our knowledge of iris recognition. Thus, the data-driven model shows good generalization performance for the data uncertainty from illumination variation.

E. Cross-sensor matching

Compared with same-sensor recognition, the cross-sensor recognition challenge is proposed for less constrained scenes. Cross-sensor recognition often occurs in the following scenarios: 1) system reconfiguration, 2) integrated systems, 3) distributed systems with multiple acquisition devices. In these scenarios, users enroll and identify using different devices, leading to device parameter variation. Thus, cross-sensor recognition not only considers representation learning for discriminative identity features but also faces the alignment of the feature spaces.



(a) Comparison on NDCS.

(b) Comparison on CSIR.

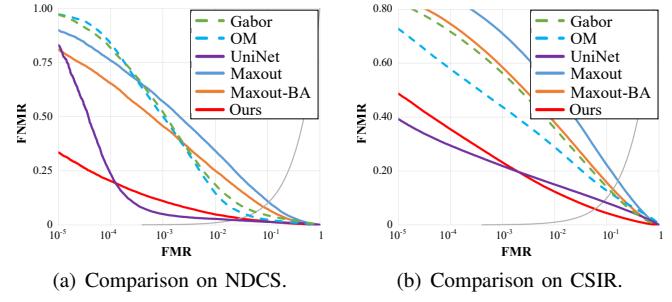
Fig. 6: DET curves of different methods on the **cross-sensor** datasets under the **within-database** protocol. The red solid lines are the results of the proposed method. (best viewed in color)

To evaluate the proposed method on cross-sensor recognition, we conduct experiments on two real-world cross-sensor datasets, ND-CrossSensor-Iris-2013 (NDCS) and CASIA Cross Sensor Iris Recognition (CSIR), under two protocols. Table IV and Table V report the detailed results of our method and compared methods under the within- and cross-database protocols, respectively. In addition, Fig. 6 and Fig. 7 plot DET curves of different approaches for a qualitative comparison.

Results under the within-database protocol. For the cross-sensor iris recognition under the within-database protocol, Table IV reports the detailed results of our method and the compared methods. We can reach the following conclusions.

1. *The proposed method shows better recognition performance than the compared methods on two large-scale cross-sensor datasets.* Specifically, CSIN was proposed in 2019 and is the state-of-the-art method for cross-sensor recognition. Our method has the lowest EER of 1.59% on the NDCS dataset, which drops by 18.87% on the relative value over CSIN. Meanwhile, our work achieves the lowest EER value of 4.52% on the CSIR dataset, a more challenging cross-sensor dataset, and drops by 34.20% on the relative value over CSIN. In terms of other error indices, the proposed method is lower than the compared methods on both cross-sensor datasets. These experimental results indicate that our method has more competitive advantages than other approaches on cross-sensor within-database recognition. In addition, compared with same-sensor iris recognition (Table II), UGCL handles hard samples with large uncertainty and significantly improves recognition performance.

2. *Compared with improvement in same-sensor recognition, our method shows a more noticeable performance improvement over the baseline for cross-sensor recognition.* According to Table IV, the proposed method drops by more than 2% on absolute values over the baseline on two datasets in terms of EER. This large decrease in EER suggests the following. 1) In cross-sensor recognition, the variation in devices plays a vital role in the degradation of recognition performance. 2) Our approach effectively alleviates the performance degradation from device variations. More specifically, UE effectively extracts the identity feature from the samples in complicated scenarios with multiple acquisition devices, thus decreasing



(a) Comparison on NDCS.

(b) Comparison on CSIR.

Fig. 7: DET curves of different methods for **cross-sensor** recognition under the **cross-database** protocol. Non-deep learning methods are marked as dashed lines while deep learning methods are marked as solid lines. The red solid lines are the results of the proposed method. (best viewed in color)

TABLE IV: Quantitative comparison of cross-sensor recognition under the **within-database** protocol (%). Note: † means the baseline method.

		FNMR@FMR	Maxout	CSIN	Maxout-BA†	Ours
NDCS	EER	5.44	1.96	3.75	1.59 (\downarrow 2.16)	
	10^1	3.41	0.84	1.81	0.74	
	10^{-1}	33.68	4.83	19.50	4.80	
	10^{-3}	69.64	12.81	52.00	15.20	
CSIR	EER	11.31	6.87	10.10	4.52 (\downarrow 5.58)	
	10^1	12.26	5.45	10.16	2.72	
	10^{-1}	51.99	25.96	46.90	17.31	
	10^{-3}	80.70	51.64	75.30	41.58	

the number of recognition failure cases.

3. *The quantitative comparison verifies the effectiveness of deep iris recognition in the cross-sensor recognition task.* Cross-sensor recognition is generally regarded as a more challenging task than same-sensor recognition. From Table II and Table IV, we can find that our method shows excellent performance close to the same-sensor recognition on the NDCS dataset and promising results on the CSIR dataset. These experimental results demonstrate that it is feasible to develop and apply deep iris learning for cross-sensor iris recognition.

Results under the cross-database protocol. According to Table V, we can obtain similar conclusions as those under the within-database protocol.

1. *Similar to the results under the within-database protocol, our method has a similar advantage in recognition performance under the cross-database protocol.* On the NDCS dataset, the proposed method has the second-lowest EER value of 2.98%, and the best result (UniNet) is only 0.74% lower than ours. However, according to [10], UniNet has its enhancement module and knows the partial acquisition condition of the NDCS dataset. Moreover, the lowest EER of 5.92% on the CSIR dataset is achieved by our method. In terms of other error indices, our method usually has lower values than the compared methods on these two datasets. The experimental results illustrate that the UE's advantage in recognition performance can be extended from a within-database scenario to a cross-database scenario.

TABLE V: Quantitative comparison on the NDCS and CSIR datasets under the **cross-database** protocol (%). Note: † means the baseline method.

FNMR@FMR		Gabor	OM	UniNet	Maxout	Maxout-BA†	Ours
NDCS	EER	5.91	4.37	2.24	10.47	4.28	2.98 (\downarrow 1.30)
	10^1	4.39	2.40	1.39	11.00	2.13	1.58
	10^{-1}	52.07	50.15	4.99	56.90	17.25	11.03
	10^{-3}	97.03	97.28	82.98	89.91	39.18	33.45
CSIR	EER	12.26	11.75	8.66	16.42	13.22	5.92 (\downarrow 7.30)
	10^1	14.08	12.69	8.21	22.24	15.77	4.40
	10^{-1}	56.27	43.79	21.85	71.02	58.47	23.19
	10^{-3}	82.56	72.77	39.25	92.59	84.86	48.72

TABLE VI: Quantitative results for ablation study on the CSIR dataset under the cross-database protocol (%).

FNMR@FMR	EER	10^1	10^{-1}	10^{-3}
Baseline	13.22	15.77	58.47	84.86
Baseline+P	9.95	9.91	37.03	61.04
Baseline+P+D	7.03	5.67	25.97	50.14
Baseline+D	8.61	7.83	33.61	57.60
Baseline+D+U	7.85	6.74	30.93	55.78
Full (+P+D+U)	5.92	4.40	23.19	48.72

2. *The proposed method achieves further improvement in recognition performance over the baseline under the cross-database protocol.* On the NDCS dataset, our method's EER is significantly lower than that of the baseline model, specifically, dropping by 1.30% on the absolute value. On the CSIR dataset, the proposed method has the lowest EER of 5.92%, which is 7.30% lower than the EER of the baseline model on the absolute value. The comparison between the proposed method and the baseline model indicates that our modules play a positive role in promoting recognition.

3. *It is becoming a trend to address cross-sensor recognition using deep iris recognition.* Log-Gabor coding and OM coding are representative methods of conventional iris recognition, while others are recently proposed deep iris recognition methods. According to Table V, deep iris recognition methods with automatic training have better recognition results than conventional recognition methods with handcrafted tuning on cross-sensor cross-database recognition. This recognition advantage would make more users prefer to adopt deep iris recognition for cross-sensor iris recognition. In this sense, we do believe that deep learning methods will be the mainstream technique of iris recognition in the future.

4. *Quantitative results under two protocols suggest that CSIR is more challenging than NDCS.* On the NDCS dataset, eye images are acquired by two types of devices, LG2200 and LG4000. These two devices are designed and produced by the same manufacturer: LG4000 upgrades LG2200 by adjusting 1) the illumination location, 2) the field of view, 3) the camera types [2]. In the CSIR dataset, eye images are acquired by IG and IK, which are produced by different manufacturers. IrisKing designs and produces the IKEMB-200, and IrisGuard designs and produces the EyeGuard AD100. Thus, the imaging gap between IG and IK is more evident than that in the NDCS dataset. This means that there exists larger data uncertainty in the CSIR dataset than in the NDCS dataset. The quantitative

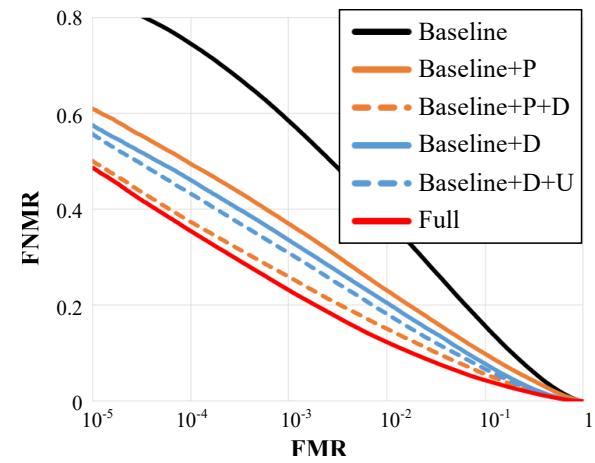


Fig. 8: DET curves of ablation models on the CSIR dataset under the cross-database protocol. The baseline model is marked as the black solid line, while the red solid lines are the results of the full model. (best viewed in color)

results in Table IV and Table V confirm this explanation.

F. Ablation study

To demonstrate each module's effectiveness, we specifically conduct an ablation study on the CSIR dataset under the cross-database protocol. In addition to the baseline and full models, we design four ablation models, as shown in Table VI (P: Enhancement module, D: UE, U: UGCL). We obtain Baseline+P+D and Baseline+D+U by removing the UGCL and the enhancement module. Baseline+P removes the UE on the basis of Baseline+P+D, while Baseline+D is generated by removing the UGCL on the basis of Baseline+D+U. The detailed results are reported in Table VI.

From Table VI, we can draw the following conclusions:

1. *Effectiveness of UE.* From Table V, the EER value of Baseline+D drops by 4.61% on the absolute value over Baseline, while the EER value of Baseline+P+D drops 2.92% on the absolute value over Baseline+P. Discarding the probabilistic model, we have to apply the deterministic point to represent the iris image. However, the deterministic point-based representation forces the feature to describe uncertainty information and fails to learn a discriminative feature. These quantitative results demonstrate the effectiveness of UE.

2. *Effectiveness of UGCL.* With the help of UGCL, the full model (Full) decreases the EER value by 1.11% compared

TABLE VII: Quantitative comparison of the proposed method on the Distance dataset using different segmentation methods. (%) Note: * is the segmentation method for the normalized iris images downloaded from the CASIA website².

Methods	SFR	mHdis	EER	FNMR@FMR		
				10^1	10^{-1}	10^{-3}
Osiris	4.26	0.88	0.92	0.10	2.91	9.87
RTV-L1	0.80	1.04	1.12	0.13	3.94	12.86
IrisParseNet	0	0.81	0.87	0.09	2.73	8.58
CASIA*	0	0.93	0.94	0.10	3.04	10.71

with Baseline+P+D, which suggests that UGCL makes the model converge to a better optimum. The recognition promotion of UGCL also serves on other ablation models, i.e., Baseline+D+U upgrades Baseline+D by introducing UGCL, and its EER value drops by 0.76%. UGCL puts the training samples in order by considering their uncertainty, which improves recognition accuracy. By employing UGCL, the EER values under both within- and cross-database protocols would significantly decrease.

3. *Effectiveness of the enhancement module.* Compared with Baseline, the ablation model (Baseline+P) achieves lower values in all error indices. A similar reduction in error indices also occurs between Baseline+D+U and Full. Without the powerful enhancement module, the input images' uncertainty sharply increases, degrading the recognition performance. These decreases illustrate the positive impact of the enhancement module on recognition.

4. *Other.* Comprehensive comparisons among ablation models indicate that each module proposed in our method can promote recognition performance. These efforts effectively alleviate the data uncertainty in iris recognition and achieve satisfying recognition performance on different datasets.

G. Segmentation effect

In an iris recognition system, segmentation errors affect the recognition performance of a recognition method. Specifically, the proposed approach achieves feature matching without segmentation masks, which is a potential advantage since this matching strategy reduces the impact of the segmentation error. However, segmentation performance has a substantial influence on the localization accuracy, hence, it influences the recognition performance.

In order to verify the impact of iris segmentation in recognition accuracy, we compare with other three segmentation techniques, including OsIris [49], RTV-L1 [50], and IrisParseNet [51], and report their performance in Table VII. We adopt the Hough circle detection [46] to obtain the iris and pupil localization of an eye image, and then achieve iris normalization using an affine transform [48]. SFR denotes segmentation failure rate, and it measures the ratio of the segmentation failure cases over all testing images. mHdis (i.e., mean normalized Hausdorff distance) [51] evaluates the localization accuracy by computing Hausdorff distance for iris

inner or outer boundary. Besides, in the recognition process, we remove 64 images that failed to be segmented so that there are the same testing pairs for a fair recognition comparison.

According to Table VII, the recognition performance of the proposed method is highly related to localization accuracy. Specifically, the ranking of localization performance is the same as the ranking of recognition performance, which demonstrates that the recognition performance of our method is highly dependent on the localization accuracy. Besides, we can obtain the best localization results according to the segmentation mask generated by IrisParseNet. The better localization results can further improve the recognition performance of our method.

H. Efficiency analysis

To avoid interference from the operation environment, we conduct multiple repeated experiments to obtain execution time. Specifically, we do not consider data loading and matching for feature extraction and only record the average value of five execution times of 100 repeated extractions (i.e., the average of 500 repeated feature extractions). For feature matching, a 1:10000 matching task (comparing the test feature with 10000 templates/vectors) is repeatedly conducted five times, and the average of the five matchings is reported as the matching time.

It should be noted that the above methods are not implemented on a single platform. Only the feature extractions of deep-learning methods are implemented in PyTorch and accelerated by GPU. The rest are performed in MATLAB without GPU acceleration. Table VIII lists the execution time and parameter size of the proposed approach and compared methods. Meanwhile, to evaluate the efficiency and effectiveness of the algorithm simultaneously, we report the product results of *EER (%)* and *total time (ms)*, named PET, in Table VIII. The product can roughly provide a comprehensive assessment of the error rate and execution time for iris recognition.

From Table VIII, we note an interesting phenomenon. Conventional iris recognition has a slight advantage in feature extraction time compared with deep iris recognition, while conventional iris recognition cannot maintain this advantage in the recognition stage. We believe that feature extraction's time cost depends on the number and size of filters (for conventional methods) or convolution kernels (for deep methods). Compared with conventional iris recognition methods with several filters, maxout-based models adopt relatively complex convolution operations. Therefore, maxout-based models spend more time on feature extraction. For matching, recent advances in deep learning for biometrics [52], [41] suggest that the hypersphere cosine distance has great potential in large-scale feature matching. We catch up with this trend and employ hypersphere cosine distance for feature matching. According to Table VIII, deep iris recognition takes much less time than conventional recognition for the 1:10000 matching task on the platform of MATLAB. Overall, deep iris recognition saves time for inference compared with conventional iris recognition according to the total time in Table VIII.

In addition to the above analysis, we think three other points should be noted.

²<http://www.cripacsir.cn/dataset/>

TABLE VIII: The quantitative comparison of efficiency analysis of different methods on the Lamp/(NDCS) datasets (%). Note: 1) Feature extraction time: average of 500 repeated extractions, Matching time: matching from 10,000 enroll templates. 2) Only the feature extractions of deep-learning methods are accelerated by **GPU**, and the rest are performed on **CPU**.

METHOD	Gabor	OM	UniNet	DGR	CSIN	Maxout	Maxout-BA†	Our
Extraction Time (ms)	3.50	0.58	7.31	41.28	6.61	2.35	2.35	2.35
Matching Time (ms)	$> 5 \times 10^3$	278.80	110.25	$> 5 \times 10^3$	0.83	0.83	0.83	0.83
Total Time (ms)	$> 5 \times 10^3$	279.38	117.55	$> 5 \times 10^3$	7.44	3.18	3.18	3.18
EER (%)	3.56	3.64	3.30	0.61	(1.96)	2.28	1.70	0.45
PET (EER×Time)	$> 5 \times 10^3$	1016.94	387.91	$> 5 \times 10^3$	(14.58)	7.25	5.41	1.43
#Params (MB)	–	–	0.13	3.19	16.45	4.10	4.10	4.10

1. *Deep iris recognition has a significant advantage in parameter tuning.* Conventional iris recognition manually designs and tunes filters based on prior knowledge. However, it is agnostic to tune the parameters for a high-accuracy model manually, which means that tuning for the conventional model is always time-consuming. In contrast, we could set a few hyperparameters for deep iris recognition, and then the machine automatically learns the model with satisfying recognition performance in a limited time. Taking our model training as an example, we train and test our method (including pretraining, fine-tuning, testing, and reporting the result) from scratch within 12 GPU hours. The automatic parameter learning of deep learning can promote the development of iris recognition.

2. *In terms of PET, our method outperforms the compared methods with a more significant improvement.* From Table VIII, we note that our model has the lowest PET of 1.43 on the Lamp dataset, which means that the proposed approach achieves a good balance between execution time and error rate.

3. *Our method achieves significant performance improvement without a noticeable increase in the number of parameters.* In the last row of Table VIII, we report the parameter size³ of deep iris recognition methods. The model size of the proposed method is almost equivalent to that of Maxout or Maxout-BA, which suggestss that the proposed method achieves the best recognition performance in a short recognition time with almost no increase in the number of parameters.

V. CONCLUSION

Iris recognition under a real-world scene is easily affected by uncertain acquisition factors, including eyes, devices, and environment. The conventional deterministic point representation for iris recognition does not consider these factors and forces systems to obtain ambiguous features (feature ambiguity dilemma). In this paper, we explore acquisition uncertainty and propose uncertainty embedding (UE) and uncertainty-guided curriculum learning (UGCL) to circumvent this dilemma. UE utilizes a probabilistic representation to encode the identity and uncertainty into the mean and variance of the distribution. The independent encoding for identity and uncertainty reduces the impact of acquisition uncertainty on identity features. UGCL selects easy features according to the data uncertainty to learn the model at the early

³The tool we take for model size calculation is from PyTorch-OpCounter. (<https://github.com/Lyken17/pytorch-OpCounter>)

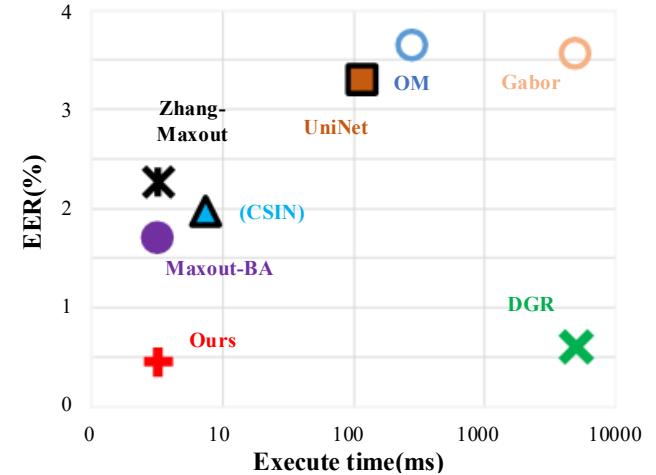


Fig. 9: A comparison of the quality and speed of the proposed method and compared methods on the Lamp/(NDCS) dataset. We visualize the EER versus the execution time. The execution time axis is in the log scale.

training stage and gradually adds hard features for training as the iteration progresses. The strategy effectively improves the recognition performance of the model via a reasonable arrangement of hard features. An instance-level enhancement module is developed to resist image noise and changing acquisition conditions. This enhancement utilizes local and global statistics to mitigate the pixel-level uncertainty for each image. Extensive experimental results on six iris benchmark datasets validate the effectiveness and generalization of the proposed method in near-infrared iris recognition. In future work, we will continue to explore deep uncertainty learning in iris recognition and extend it to more iris-related tasks.

ACKNOWLEDGEMENTS

We thank the associate editor and the reviewers for their useful feedback that improved this paper. This work is supported by National Natural Science Foundation of China (Grant No. U1836217, U21B2045, 62176025, 61622310, 62071468), Beijing Natural Science Foundation (Grant No. JQ18017), Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA27040700), Youth Innovation Promotion Association CAS (Grant No. 2015109), and sponsored by CAAI-Huawei MindSpore Open Fund.

REFERENCES

- [1] Y. He, J. Cui, T. Tan, and Y. Wang, "Key techniques and methods for imaging iris in focus," in *IEEE International Conference on Pattern Recognition*, vol. 4, 2006, pp. 557–561.
- [2] K. W. Bowyer, S. E. Baker, A. Hentz, K. Hollingsworth, T. Peters, and P. J. Flynn, "Factors that degrade the match distribution in iris biometrics," *Identity in the Information Society*, vol. 2, no. 3, pp. 327–343, 2009.
- [3] P. R. Nalla and A. Kumar, "Toward more accurate iris recognition using cross-spectral matching," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 208–221, 2016.
- [4] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34–45, 2019.
- [5] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in Neural Information Processing Systems*, 2017, pp. 5574–5584.
- [6] T. Yu, D. Li, Y. Yang, T. M. Hospedales, and T. Xiang, "Robust person re-identification by modelling feature uncertainty," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 552–561.
- [7] Y. Shi and A. K. Jain, "Probabilistic face embeddings," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 6902–6911.
- [8] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.
- [9] Z. Sun and T. Tan, "Ordinal measures for iris recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2211–2226, 2008.
- [10] Z. Zhao and A. Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3809–3818.
- [11] Q. Zhang, H. Li, Z. Sun, and T. Tan, "Deep feature fusion for iris and periocular biometrics on mobile devices," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2897–2912, 2018.
- [12] Z. Sun, L. Wang, and T. Tang, "Ordinal feature selection for iris and palmprint recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3922–3934, 2014.
- [13] A. Gangwar and A. Joshi, "Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in *Proceedings of IEEE International Conference on Image Processing*, 2016, pp. 2301–2305.
- [14] S. Isobe and S. Arai, "Deep convolutional encoder-decoder network with model uncertainty for semantic segmentation," in *IEEE International Conference on INnovations in Intelligent SysTems and Applications*, 2017, pp. 365–370.
- [15] J. Choi, D. Chun, H. Kim, and H.-J. Lee, "Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 502–511.
- [16] Y. Gal, "Uncertainty in deep learning," *University of Cambridge*, vol. 1, no. 3, 2016.
- [17] M. E. Paté-Cornell, "Uncertainties in risk analysis: Six levels of treatment," *Reliability Engineering and System Safety*, vol. 54, no. 2-3, pp. 95–111, 1996.
- [18] C. Belcher and Y. Du, "A selective feature information approach for iris image-quality measure," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 572–577, 2008.
- [19] X. Li, Z. Sun, and T. Tan, "Comprehensive assessment of iris image quality," in *IEEE International Conference on Image Processing*, 2011, pp. 3117–3120.
- [20] Y. Wang, W. Gan, J. Yang, W. Wu, and J. Yan, "Dynamic curriculum learning for imbalanced data classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 5017–5026.
- [21] M. Sachan and E. Xing, "Easy questions first? a case study on curriculum learning for question answering," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2016, pp. 453–463.
- [22] D. Yambay, B. Becker, N. Kohli, D. Yadav, A. Czajka, K. W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore *et al.*, "Livdet iris 2017—iris liveness detection competition 2017," in *IEEE International Joint Conference on Biometrics*, 2017, pp. 733–741.
- [23] K. Wang and A. Kumar, "Toward more accurate iris recognition using dilated residual features," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3233–3245, 2019.
- [24] N. Liu, J. Liu, Z. Sun, and T. Tan, "A code-level approach to heterogeneous iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 10, pp. 2373–2386, 2017.
- [25] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, p. 21, 2004.
- [26] C.-W. Tan and A. Kumar, "Accurate iris recognition at a distance using stabilized iris encoding and zernike moments phase features," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3962–3974, 2014.
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [28] F. Wang, X. Xiang, J. Cheng, and A. L. Yuille, "Normface: L2 hypersphere embedding for face verification," in *Proceedings of the ACM International Conference on Multimedia*, 2017, pp. 1041–1049.
- [29] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 212–220.
- [30] X. Wu, R. He, Z. Sun, and T. Tan, "A light cnn for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018.
- [31] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan, "Iris recognition with off-the-shelf cnn features: A deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2017.
- [32] K. Nguyen, C. Fookes, and S. Sridharan, "Constrained design of deep iris networks," *IEEE Transactions on Image Processing*, vol. 29, pp. 7166–7175, 2020.
- [33] M. Ren, Y. Wang, Z. Sun, and T. Tan, "Dynamic graph representation for partially occluded biometrics," in *AAAI Conference on Artificial Intelligence*, 2020, pp. 2301–2305.
- [34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.
- [35] J. Wei, Y. Wang, X. Wu, Z. He, R. He, and Z. Sun, "Cross-sensor iris recognition using adversarial strategy and sensor-specific information," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2019.
- [36] J. Postels, F. Ferroni, H. Coskun, N. Navab, and F. Tombari, "Sampling-free epistemic uncertainty estimation using approximated variance propagation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2931–2940.
- [37] U. Zafar, M. Ghafoor, T. Zia, G. Ahmed, A. Latif, K. R. Malik, and A. M. Sharif, "Face recognition with bayesian convolutional networks for robust surveillance systems," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, p. 10, 2019.
- [38] S. Khan, M. Hayat, S. W. Zamir, J. Shen, and L. Shao, "Striking the right balance with uncertainty," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 103–112.
- [39] J. Chang, Z. Lan, C. Cheng, and Y. Wei, "Data uncertainty learning in face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5710–5719.
- [40] A. C. Miller, N. J. Foti, and R. P. Adams, "Variational boosting: Iteratively refining posterior approximations," in *International Conference on Machine Learning*, 2017, pp. 2420–2429.
- [41] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.
- [42] L. Xiao, Z. Sun, R. He, and T. Tan, "Coupled feature selection for cross-sensor iris recognition," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2013, pp. 1–6.
- [43] J. K. Pillai, M. Puertas, and R. Chellappa, "Cross-sensor iris recognition through kernel learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 73–85, 2013.
- [44] CASIA-iris-V4 Data Set. [Online]. Available: <http://www.cripacsir.cn/dataset/casia-irisv4/>
- [45] ND-Cross-Sensor-Iris-2013 Data Set. [Online]. Available: <https://cvrl.nd.edu/projects/data/\#nd-crosssensor-iris-2013-data-set>
- [46] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1670–1684, 2008.
- [47] C.-W. Tan and A. Kumar, "Unified framework for automated iris segmentation using distantly acquired face images," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4068–4079, 2012.

- [48] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Efficient iris recognition by characterizing key local variations," *IEEE Transactions on Image Processing*, vol. 13, no. 6, pp. 739–750, 2004.
- [49] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "Osiris: An open source iris recognition software," *Pattern Recognition Letters*, vol. 82, pp. 124–131, 2016.
- [50] Z. Zhao and K. Ajay, "An accurate iris segmentation framework under relaxed imaging constraints using total variation model," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3828–3836.
- [51] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun, "Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition," *IEEE Transactions on information forensics and security*, vol. 15, pp. 2944–2959, 2020.
- [52] S. Li, J. Xu, X. Xu, P. Shen, S. Li, and B. Hooi, "Spherical confidence learning for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 629–15 637.



Zhenan Sun is currently a Professor in CRIPAC, NLPR, CASIA, China. He received the B.E. degree in industrial automation from Dalian University of Technology, and the M.S. degree in system engineering from Huazhong University of Science and Technology, and the Ph.D. degree in pattern recognition and intelligent systems from CASIA in 1999, 2002, and 2006 respectively. His research focuses on biometrics and pattern recognition. He is a fellow of IEEE IAPR. He is an Associated Editor of IEEE on Biometrics, Behavior, and Identity Science.



Jianze Wei received the B.E. and M.S. degrees in communication engineering from the Civil Aviation University of China, Tianjin, China, in 2015 and 2018, respectively. He is currently studying for the Ph.D. degree in University of Chinese Academy of Sciences, Beijing, China. His research interests include biometrics, machine learning and computer vision.



Huaibo Huang is currently an Associate Professor in CRIPAC, NLPR, CASIA, China. He received the B.E. degree in Measurement and Control Technology and Instrument from Xi'an Jiaotong University in 2012, the M.E. degree in Optical Engineering from Beihang University in 2016, the the Ph.D. degree in Pattern Recognition and Intelligent System from CASIA in 2019. His current research interests include computer vision and pattern recognition.



Yunlong Wang is currently an Associate Professor with Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA) , China. He received the B.E. degree and the Ph.D. degree in Department of Automation, University of Science and Technology of China. His research focuses on pattern recognition, machine learning, light field photography and biometrics.



Ran He received the B.E. and M.S. degrees in computer science from the Dalian University of Technology, in 2001 and 2004, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the Chinese Academy of Sciences in 2009. Since 2010, he has been a Full Professor with the National Laboratory of Pattern Recognition (NLPR). His research interests focus on information theoretic learning, pattern recognition, and computer vision. He serves as an Associate Editor of *Neurocomputing* (Elsevier) and serves on the program committee of several conferences.